Linear Regression ‘Selection’ & Inference

STAT 245

If INFERENCE is your goal...

CIs Beyond Prediction Plots

Slope coefficient of 0 (or CI crossing 0) means:

```
my_model <- lm(response ~ pred1 + pred2, data = my_data)
summary(my_model)
```

```
##
## Call:
## lm(formula = response ~ pred1 + pred2, data = my_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.75871 -0.73156 -0.06018  1.04880  1.85244
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.9823     1.6326   1.214  0.25011
## pred1        -0.7523     0.2419  -3.110  0.00992 **
## pred2B        1.5970     0.9752   1.638  0.12976
## pred2C        0.4770     0.9913   0.481  0.63980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

CIs Beyond Prediction Plots

Slope coefficient of 0 (or CI crossing 0) means:

```
my_model <- lm(response ~ pred1 + pred2,
data = my_data)
confint(my_model)
```

```
##                   2.5 %     97.5 %
## (Intercept) -1.6111048  5.5756121
## pred1       -1.2845856 -0.2199575
## pred2B      -0.5494036  3.7433632
## pred2C      -1.7048502  2.6589223
```

Hypothesis Tests

May use but see later IC slides for alternative

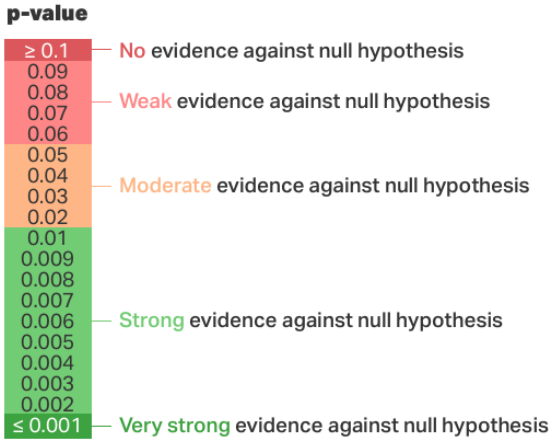
```
car::Anova(my_model)
```

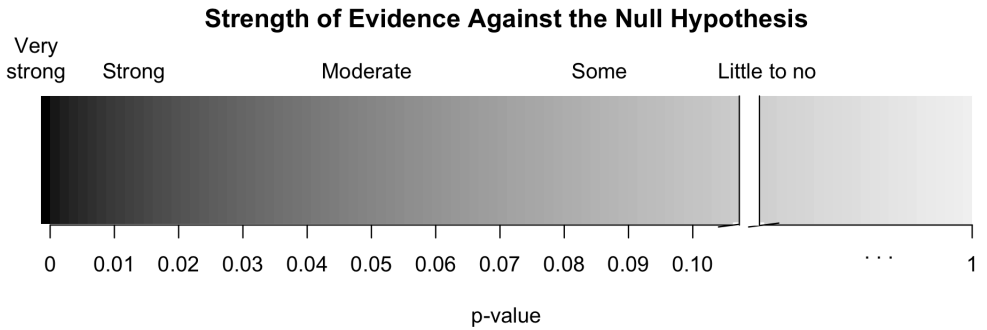
```
## Anova Table (Type II tests)
##
## Response: response
##            Sum Sq Df F value   Pr(>F)
## pred1     23.0010  1  9.6749 0.009917 **
## pred2      6.7005  2  1.4092 0.285193
## Residuals 26.1512 11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01
'*' 0.05 '.' 0.1 ' ' 1
```

~~`pred1` is significant~~

no no nopeity no no never!

Judge *your* strength of evidence (~~significance threshold~~)





Images: left, analythical.com; right, MT State intro stat

~~`pred1` is significant (p-value < 0.05)~~

There is strong evidence (p-value = 0.0099) that `pred1` is associated with `response`, with the expected `response` increasing by about ____ (95% CI: ____ to ____) when `pred1` increases by 1 (if `pred2` doesn't change).

To present `Anova()` results, use your `summary()`, equation, and ESPECIALLY prediction plot to tell the story in context!

Data

BECHDEL TEST HOW THEY STACK UP	TWO FEMALE CHARACTERS	FEMALE CONVERSATION	NOT ABOUT MEN
The Social Network	✓	✗	✗
Mad Max Fury Road	✓	✓	✓
Finding Nemo	✓	✗	✗
Jack Reacher	✓	✗	✗
The Hunger Games	✓	✓	✓
Toy Story	✓	✗	✗
Up	✗	✗	✗
The Fast and the Furious	✓	✗	✗
Silence of the Lambs	✓	✓	✓
Shawshank Redemption	✗	✗	✗

Bechdel Test Model

What is WRONG here?

```
movie_mod <- lm(roi ~ clean_test +
                  budget_2013 +
                  title,
                data = bechdel_13)
```

Better-planned Model

```
bechdel_mod <- lm(roi ~ clean_test +
                    budget_2013,
                  data = bechdel_13)
coef(bechdel_mod)
```

```
##       (Intercept) clean_testdubious
clean_testmen  clean_testnotalk
##        6.89076912         2.71983184
1.17133411         0.04824852
## clean_testnowomen        budget_2013
##        1.18334508        -0.03661270
```

Selection w/ Information Criteria

Why ICs? Balance conflicting goals

- Want model that *fits the data as well as possible* (which pushes us toward more predictors)
- Want *truth, efficiency* (correctly judging when predictors aren't associated with the response).
- Solution: **minimize** the quantity
  $-(2ln(\mathcal{L}) - \text{penalty}) = -2ln(\mathcal{L}) + \text{penalty}$

AIC Defined

- $\mathcal{L}$ is the "Likelihood" of data given model
- In R: `logLik(model)` gives $ln(\mathcal{L})$
- AIC: $-2ln(\mathcal{L}) + 2k$
  - k is the number of coefficients being estimated (don't forget σ!)
  - **Smaller AIC is better.**
- In R: `AIC(model)` or `AIC(model1, model2, model3 ... )`

BIC Defined

- BIC: $-2ln(\mathcal{L}) + ln(n)k$
  - n is the number of rows in dataset
  - k is the number of coefficients estimated.
  - **Smaller BIC is better.**
- In R: `BIC(fitted_model)` or `BIC(model1, model2, model3 ... )`

Comprehension Check: Verify BIC

```
coef(bechdel_mod)
```

```
##       (Intercept) clean_testdubious     clean_testmen  clean_testnotalk
##        6.89076912        2.71983184        1.17133411        0.04824852
## clean_testnowomen       budget_2013
##        1.18334508       -0.03661270
```

```
nrow(bechdel_13)
```

```
## [1] 1600
```

```
logLik(bechdel_mod)
```

```
## 'log Lik.' -6962.028 (df=7)
```

Verify that the BIC for this model is 13976.

Decisions with ICs

Rules of thumb (not laws) for decisions with ICs

- IC *lower by at least 3 units* = notably better
- If 2+ models have ICs *within* 3 IC units of each other, they fit about same
- If $\Delta IC < 3$ or so: model with smallest k (fewest predictors) is better

All-subsets selection

Avoid in favor of interpreting the full model or comparing a few key models if possible

- Use `dredge()` function (`MuMIn` package) to get and display ICs for every possible combo of predictors.
- First ensure dataset has no missing values,
- then set `na.action = 'na.fail'` input for our model (with `update()` or in `lm()`).

`dredge()` example

```
require(MuMIn)
bechdel_mod <- bechdel_mod |>
  update(na.action = 'na.fail')
bechdel_dredge <- dredge(bechdel_mod,
                         rank = 'BIC')
bechdel_dredge
```

`dredge()` example

What is the best model according to BIC, for this dataset?

```
## Global model call: lm(formula = roi ~ clean_test + budget_2013, data =
bechdel_13,
##     na.action = "na.fail")
## ---
## Model selection table
##   (Int) bdg_2013 cln_tst df    logLik     BIC delta weight
## 2 7.282 -0.03589          3 -6963.452 13949.0  0.00  0.996
## 1 5.220                   2 -6972.569 13959.9 10.86  0.004
## 4 6.891 -0.03661       +  7 -6962.028 13975.7 26.66  0.000
## 3 5.157                +  6 -6971.235 13986.7 37.70  0.000
## Models ranked by BIC(x)
```

Which IC to use?

- AIC and BIC (and `Anova()`) may yield different conclusions, especially if the dataset is large.
- Recommendation: choose **one** to use *a priori* (before making calculations).
- Prefer BIC to be "more conservative" (larger penalty means predictor must improve fit *more* before we judge it worthwhile)

Quantities derived from (A)IC

- ΔAIC is the AIC for a given model, minus the AIC of the best one in the dataset. (Same for ΔBIC)
- *Akaike* or *BIC weights* are values (ranging from 0-1) that measure the weight of evidence suggesting that a model is the best one (given that there is one best one in the set)

Important Caution

- **Very important**: IC can **ONLY** be compared for models...
  - With the exact same response variable...
  - and the exact same rows of data.

Conclusions

Evidence of association?

Evidence the arrow (in causal diagram) is strongly, detectably *there*?

```
no_bechdel_mod <- lm(roi ~ budget_2013,
                     data = bechdel_13)
AIC(bechdel_mod, no_bechdel_mod)
```

```
##                  df      AIC
## bechdel_mod       7 13938.06
## no_bechdel_mod    3 13932.90
```

~~`pred1`~~ **~~is significant~~**

There is strong evidence (AIC lower by about 5) that a movie's return on investment is associated with its Bechdel test score. Specifically, after controlling for any effect of budget, model results suggest that...[use pred plot here to say which test scores predict more/less ROI!].

To present IC results, use your `summary()`, equation, and ESPECIALLY prediction plot to tell the story in context!